

Standardy i dobre praktyki budowania kolekcji cyfrowych

Abstrakt

Autorka przedstawia główne zasady tworzenia poprawnej kolekcji cyfrowej określone w dokumencie amerykańskiej organizacji NISO (National Information Standards Organization) pt.: A Framework of Guidance for Building Good Digital Collections. Dotyczą one elementów planowania projektowego i zarządzania kolekcją, zasobów, obiektów i metadanych. Jest to próba przyswojenia amerykańskich i europejskich dobrych praktyk dla polskiej pragmatyki zawodowej. Tworząc kolekcję cyfrową należy stosować podobne podejście do zasad, które zostały opisane poniżej.

WSTĘP

Światowe zasoby elektroniczne organizowane w duże kolekcje powstają mniej więcej od 1995 roku i jest w tej chwili bardzo dużo zarówno tych komercyjnych, jak i budowanych przez instytucje publiczne czy fundacje. Ich różnorodność jest ogromna od fachowych po amatorskie, od wąskich tematycznie przez uniwersalne, od prowadzonych zgodnie ze standardami międzynarodowymi do odbiegających od wszelkich zasad.

Budowanie kolekcji cyfrowych jest dość dobrze rozpoznany zadaniem i opisanym w miarę szczegółowo, nie tylko w zagranicznych opracowaniach, ale także w Polsce. Biblioteka Narodowa, która jest od niedawna Narodowym Centrum Kompetencji w zakresie digitalizacji ma za zadanie świadczyć pomoc wszystkim, którzy chcą się podjąć cyfryzacji. W roku 2010 ukazał się podręcznik Biblioteki Narodowej zatytułowany „Digitalizacja piśmiennictwa”¹, ponadto odbywają się w niej praktyczne szkolenia, warsztaty i seminaria powiązane z różnymi aspektami digitalizacji, także prawnymi. W roku 2008 dzięki pracy Zespołu ds. digitalizacji pracującego przy MKiDN opracowano także pierwsze w Polsce *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego* pod redakcją i we współpracy Grzegorza Płoszajskiego, Tomasza Kaloty, Dariusza Paradowskiego oraz Kazimierza Schmidta. Opublikowano je ze wsparciem Politechniki Warszawskiej: <http://bcpw.bg.pw.edu.pl/dlibra/docmetadata?id=1262>.

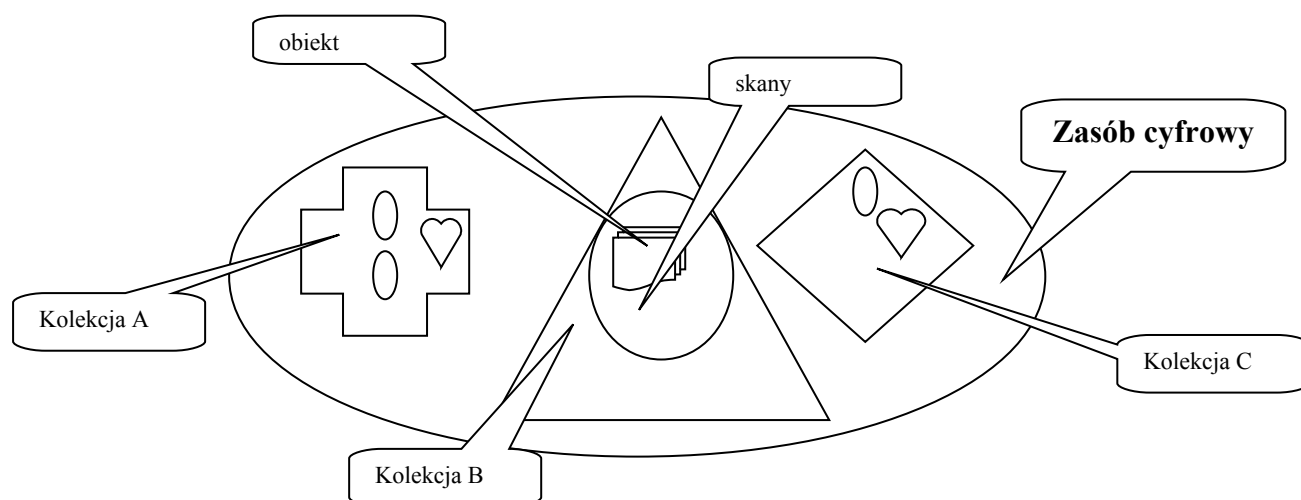
Ponadto mamy w kraju kilka ośrodków, które mogą służyć radą i konsultacją, ponieważ mają długoletnie doświadczenie zawodowe w tym zakresie. Są to: Uniwersytet Mikołaja Kopernika w Toruniu, Uniwersytet Wrocławski, Uniwersytet Poznański, Zakład Narodowy Ossolineum, Biblioteka Kórnicka czy Biblioteka Śląska. Znajdują się w nich pracownie digitalizacji i zespoły fachowców, którzy poradzą, jak zbudować poprawnie kolekcje, jak je opisać czy nimi zarządzać. Mamy także polskie oprogramowanie dLibra i dMuseion dla obsługi tego typu zadań i istotne wsparcie technologiczne instytucji naukowej PAN — Poznańskiego Centrum Superkomputerowo-Sieciowego. Zespół młodych informatyków tego ośrodka prowadzi bardzo ambitne programy rozwojowe, dzięki którym powstała Federacja Polskich Bibliotek Cyfrowych skupiająca w roku 2010 ponad 60 odrębnych bibliotek cyfrowych gromadzących pół miliona obiektów i udostępnionych od 2009 roku w Europeanie.

¹ *Digitalizacja piśmiennictwa* pod red. Dariusza Paradowskiego, Biblioteka Narodowa Warszawa 2010. ISBN 978-83-7009-706-6.

Skala digitalizacji w Polsce nie jest duża, ale stale się zmienia. W 2006 roku mieliśmy zaledwie kilka bibliotek cyfrowych, w 2008 już tworzyliśmy niewielką sieć 17 z nich, która udostępniła online ponad 120 tysięcy obiektów. Był to dość dynamiczny proces pokazujący i ambicje bibliotekarzy i przekonanie, że zadanie to jest standardowe dla bibliotek i innych instytucji posiadających ważne dla Polski kolekcje. Można powiedzieć, że na owe lata było to niemałe osiągnięcie, zważywszy na to, że instytucje państwowe dotujące kulturę czy naukę włączyły się finansowo do cyfryzacji dopiero w roku 2006, wcześniej mieliśmy do czynienia z oddolnymi inicjatywami.

Dlaczego standardy są ważne?

Wartościowe kolekcje, jeśli mają przetrwać kolejne rewolucje technologiczne, stanowić wspólne europejskie dziedzictwo, muszą oprzeć się o kilka zasad, pozwalających stworzyć profesjonalne archiwum cyfrowe² z zapewnioną interoperacyjnością, stabilnością, możliwością powtórznego użycia i weryfikacji zasobów, zabezpieczenia obiektów oraz zgodnością z prawem autorskim czy prawem o ochronie danych osobowych. Archiwa cyfrowe (AC) nie powinny być tymczasowymi, przypadkowymi projektami poszczególnych instytucji, działającymi na niesprawdzonym oprogramowaniu, z byle jak opisanymi obiektami, ale muszą być budowane tak, by nie trzeba było powtarzać kosztownego procesu skanowania czy opisywania kolekcji.



Schemat elementów składowych zasobu cyfrowego KPBC: zasób (KPBC), kolekcje (Baltica, Pomeranica), obiekty (książka, czasopismo) i skan (1 strona z książki).

Dziś z perspektywy czasu warto już pokusić się o dokładną analizę na ile zbudowane w Polsce kolekcje zachowują standardy, jakie powinny być zachowane, by przetrwały 100 lat. Bez takiej analizy ciężko nam będzie naprawić błędy i wejść na ścieżkę naprawy. Brak koordynacji wielu działań sprawia, że nadal mamy kłopoty z jednolitym stosowaniem metadanych opisowych, nie rozumiemy prawa autorskiego w związku z tym źle opisujemy status prawy obiektu cyfrowego, dramatycznie brakuje nam wiedzy z zakresu konserwacji i zachowaniu cyfrowego zasobu. W Polsce nie ma profesjonalnych ośrodków kształcenia w tym konkretnym zakresie.

² Będą stosowała kilka terminów na określenie zasobów cyfrowych, wszystkie można traktować zamiennie, choć ich szczegółowe znaczenie nieco się różni, a zatem będziemy mówić o: *archiwum cyfrowym*, *bibliotece cyfrowej*, *kolekcji cyfrowej*, *zasobie cyfrowym*, można także użyć określenia: *archiwum elektroniczne*, ale to raczej podkreśla techniczny charakter tegoż. *Zasób cyfrowy*, *kolekcja cyfrowa* rozumiane tu są jako zbiory obiektów cyfrowych (zdigitalizowanych, elektronicznych) i metadanych (informacji tekstowych), które je opisują. *Archiwum cyfrowe* i *bibliotekę cyfrową* można rozumieć szerzej - razem z platformą technologiczną (oprogramowaniem), na której zasób i metadane są posadowione.

Wiedza, jaką mamy, wynika głównie z doświadczeń praktycznych oraz literatury zagranicznej. To za mało, by podjąć się tak trudnych wyzwań.

Mamy chyba jednak świadomość, że wszystkie procesy związane z digitalizacją powinny respektować międzynarodowe standardy tak, by nasze zasoby stały się częścią światowych kolekcji, dały się wyszukiwać, rozpoznawać, by były wiarygodne i wreszcie trwale przechowywane. Zachowanie dla przyszłości obiektów cyfrowych jest jednym z ważniejszych zadań kultury, nauki czy przemysłu mediów. Powtórzmy zatem te najważniejsze zasady, bez których nie da się zbudować poprawnej biblioteki cyfrowej i zastanówmy się, na ile ich przestrzegamy. Zasady o których będzie mowa w następnej części tego opracowania są określone w dokumencie amerykańskiej organizacji NISO (National Information Standards Organization) zatytułowanym: *A Framework of Guidance for Building Good Digital Collections*: <http://framework.niso.org/node/5>.

PLANOWANIE I PROJEKTOWANIE

Zasada 1: Dobre archiwum cyfrowe musi mieć plan działania od samego początku jego tworzenia.

Planowanie jest najistotniejszą sprawą, jeśli chcemy wyprodukować zasób dobrej jakości. Na początek należy określić grupę docelową, dla której chcemy stworzyć kolekcję i jej potrzeby w tym zakresie. Następnie należy napisać szczegółowy plan działania, który zawiera istotne elementy takie jak: cel działania, polityczne priorytety, czas trwania projektu, rodzaj kolekcji, jaki obejmuje, selekcja, proces digitalizacji, sprawy praw autorskich, dostęp, zarządzanie zasobami, technologie, rozpowszechnianie i ocena jakości. Dopiero z tak opracowanym planem można przystąpić do poszukiwania pieniędzy i realizacji projektu.

Zasada 2: Dobry pomysł potrzebuje fachowców i ekspertów.

Rola zespołu, który będzie projektem zarządzał jest bardzo istotna, od wiedzy i doświadczenia ludzi, ich zapału zależy sukces projektu. Do projektu powinno się także zatrudniać ekspertów, którym zleca się szczególne analizy, np. bibliotekarskie, archiwalne, ekonomiczne czy prawnicze, po to, by samemu nie rozwiązywać problemów, na których się nie znamy. Daje to dużą oszczędność czasu i kosztów. Prace organizacyjne i koordynacyjne z takim zespołem powinny być także prowadzone profesjonalnie z wykorzystaniem wszystkich zasobów ludzkich i zapału, jaki ludzie mają na początku tworzenia. Możemy wybrać trzy strategie doboru ludzi: ze swojej własnej instytucji, zatrudnić firmę zewnętrzną do wykonania zadania (outsourcing) lub podjąć współpracę z innym partnerem i wykorzystać jego wiedzę oraz doświadczenie. Każda ze strategii ma wady i zalety.

Zasada 3: Dobra inicjatywa na rzecz budowania biblioteki cyfrowej (BC) powinna korzystać z dobrych praktyk z zakresu zarządzania projektem.

Zarządzanie projektem jest sztuką, która ma swoją tradycję i jest szeroko opisana, szczególnie w praktyce biznesowej. Warto czerpać z tych doświadczeń. Metody opracowywania projektów są różne, ale są pewne stałe komponenty, które trzeba uwzględnić w swoim projekcie.

Planowanie:

- a) opisujemy cele i zadania oraz punkt, jaki chcemy osiągnąć kończąc projekt;
- b) określamy wskaźniki oceny projektu, które musimy osiągnąć;
- c) określamy potrzeby, jakie powinny być spełnione, by zadanie się udało;
- d) szacujemy czas potrzebny na wykonanie zadania;
- e) opisujemy szczegółowe zadania i rozpisujemy odpowiedzialność dla każdego z nich;
- f) piszemy harmonogram projektu.

Implementacja:

- a) kiedy zaczniemy realizować projekt musimy na bieżąco śledzić zadania i ich wykonanie;
- b) regularnie oceniamy postępy prac i modyfikujemy je, jeśli trzeba;
- c) gdyby pojawiły się istotne przeszkody, to zmieniamy plan.

Ocena wykonania zadania:

- a) sprawdzamy, czy osiągnięte zostały główne cele;
- b) sprawdzamy założone w planie wskaźniki;
- c) identyfikujemy zmiany;
- d) podsumowujemy zadania i wyciągamy wnioski z jego realizacji, szczególnie badamy te elementy, które nam nie wyszły;

Modyfikacja:

- e) poprawiamy błędne założenia;
- f) wprowadzamy zmiany do procesów pracy;
- g) robimy monitoring w określonym czasie.

Zasada 4: Dobry projekt posiada plan oceny.

Ocena powinna wynikać z realizacji wskaźników, które określi się podczas planowania. Wskaźniki oceny w przypadku zasobu cyfrowego można określić następująco:

1. liczba kolekcji zbudowanych w ramach jednego zasobu;
2. liczba obiektów zeskanowanych (obiektem jest tu np. 100 stronicowa książka) w ramach poszczególnych kolekcji;
3. liczba skanów wykonanych (skan jest to jedna strona książki) dla wszystkich kolekcji;
4. liczba potencjalnych użytkowników;
5. planowany wzrost liczby użytkowników po implementacji archiwum cyfrowego w %.

Ocena może skupiać się na procesach pracy zespołu i wynikach, jakie się uzyska w określonym czasie. W ocenie można też uwzględnić: zarządzanie zespołem, zasobami oraz procedury działania, jakość produktu. Ważne są wyniki: ile zeskanowano stron, ile to jest obiektów, czy działają wdrożone systemy oraz czy sprzęt i wyposażenie zostały zakupione zgodnie z założeniami i są optymalnie wykorzystane, czy liczba użytkowników biblioteki cyfrowej rośnie, itp. Metody oceny mogą być różne, np. badania grup docelowych, robienie wywiadów, badania logów transakcji lub studia przypadków. Wyniki ocen przedstawia się decydentom, donatorom, pomagają nam one także w rozwijaniu projektu.

Zasada 5: Dobry projekt jest promowany, szeroko rozpowszechnia się informację o pracach projektowych i jego realizacji.

Realizatorzy AC powinni zrobić wszystko, żeby dobrze opracowana dokumentacja projektowa była przejrzysta i dostępna dla każdego, zwłaszcza tam, gdzie projekt realizowany jest z pieniędzy podatników. Jeśli projekt może być modelowym powinno się opublikować podręcznik dobrych praktyk, by każdy mógł się z nimi zapoznać i omijać błędy, które były naszym udziałem. Sprawozdania z realizacji projektu powinny być publiczne dostępne. Ten zwyczaj szerokiego dzielenia się wiedzą praktyczną ma wpływ na rozwój i mnożenie się podobnych realizacji w poszczególnych instytucjach. Jeżeli projekt ma znaczenie lokalne, regionalne lub narodowe, powinien być opisane w publikacjach, prezentacjach, anonsowany w mediach. Skuteczne promowanie kolekcji cyfrowej przyniesie pożądane rezultaty, wzrost zainteresowania i wykorzystania. Nowoczesny plan marketingowy zakłada, że promocja projektu podąży tam, gdzie są jego potencjalni użytkownicy.

Zasada 6: Dobry projekt dotyczy całego cyklu życia kolekcji cyfrowej i związanych z nią usług.

Istotnym jest, by przy planowaniu pamiętać o tym, że projekt ma przetrwać nie tylko do zakończenia harmonogramu związanego z wnioskowaniem o pieniądze, ale i dłużej. W przypadku biblioteki cyfrowej ważnym jest, by dokonać dobrego wyboru sprzętu, platformy cyfrowej, zapewnić dalsze finansowanie, opracować taką organizację pracy zespołu, żeby po roku zespół się nie rozpadł. Należy także zaprojektować wykorzystywanie zasobów w długim okresie i niezbędne aktualizacje zasobu i stron www. Technologie się zmieniają, musimy założyć ich nieustanne modyfikacje i odnawianie. Dziś niezbędnym jest także plan zabezpieczenia kolekcji dla przyszłości. Należy obserwować trendy światowe, standardy, patrzeć, w jakim kierunku zmierzają inne podobne inicjatywy i „być zawsze na fali”. Współorganizować szersze inicjatywy, narodowe i międzynarodowe, by wspólnie zmieniać to, co stworzy się z wielkim trudem.

ZASOBY - kolekcja

Zasada 1: Dobry zasób cyfrowy jest tworzony na podstawie jasno określonej polityki rozwoju zasobu, uzgodnionej i udokumentowanej przed rozpoczęciem digitalizacji.

Każda biblioteka cyfrowa jest inna, jej rozwój zależy od polityki instytucji, która ją tworzy, jej misji i określonej grupy użytkowników. Warto jednak mieć także na uwadze użytkowników potencjalnych, o których na początku nic nie wiemy, a którzy z pewnością się pojawią. Nie działamy w społeczności lokalnej, ale w sieci globalnej, taka jest różnica. Jeśli instytucja, która tworzy archiwum cyfrowe, ma swoje trwałe, stare kolekcje, to jej cyfrowy zasób powinien być zgodny z gromadzeniem kolekcji tradycyjnej. Zasada ta nie obowiązuje przy masowej digitalizacji i przy digitalizacji na życzenie. Kryteria selekcji materiału do digitalizacji powinny być opisane i publicznie dostępne. Politykę można spisać w paru punktach na jednej kartce. Zob. politykę KPBC: <http://kpbc.umk.pl/dlibra/text?id=polityka>.

Zasada 2: Zasoby powinny być opisane w sposób pozwalający użytkownikowi uzyskać charakterystyki zasobów, w tym ich zakres, ograniczenia dostępu, własność i inne informacje istotne dla określenia autentyczności, integralności i interpretacji zasobów.

Opis kolekcji powinien pozwalać użytkownikowi zrozumieć, z jakimi zasobami i obiektami ma do czynienia (zakres i zasięg), jakim ograniczeniom związanym z wykorzystaniem podlega kolekcja, do kogo należy, jaki jest jej historyczny kontekst czy pochodzenie, w jakim celu i dla kogo została stworzona itp. Należy zadbać o podanie wszelkich kontaktów, by czytelnik mógł dopytać o status kolekcji i prawa własności. Powinno się wyjaśnić użytkownikom, jaki software jest potrzebny, by dobrze wykorzystać zasoby. Zob. opis projektu KPBC: <http://kpbc.umk.pl/dlibra/text?id=library-desc>.

Zasada 3: Nad kolekcją cyfrową trzeba stale sprawować pieczę, dobrze nią zarządzać w trakcie całego cyklu życia i dbać o jej zabezpieczenie.

Zarządzanie kolekcją polega na stałym jej redagowaniu, meliorowaniu, dodawaniu nowych obiektów, usuwaniu zbędnych, czuwaniu nad pozyskiwaniem praw autorskich, archiwizowaniu, linkowaniu do podobnych obiektów oraz zabezpieczaniu całego zasobu na płytach lub serwerach (zaleca się to drugie rozwiązanie). Zabezpieczenie to także budowanie polityki przyszłej migracji obiektów i emulacji softwaru oraz hardware'u. Warto wiedzieć, że w Polsce powstaje Krajowy Magazyn Danych, który będzie zabezpieczał wartościowe zasoby: zob. strony PCSS <https://kmd.pcss.pl/>.



Rozmieszczenie węzłów KMD
Rys. ze stron PCSS

Zasada 4: Dobry zasób jest szeroko dostępny bez zbędnych przeszkód w jego użytkowaniu. Zasoby powinny być dostępne dla osób niepełnosprawnych i efektywnie wykorzystywane w połączeniu z technologiami adaptacyjnymi.

Dostępność oznacza, że kolekcja jest w sieci 24/h, co związane jest z implementacją odpowiedniego systemu zabezpieczenia i administrowania. Ktoś musi stale czuwać na bibliotekę cyfrową i reagować na wszystkie techniczne problemy. Dostępność nie oznacza, że wszystkie materiały są za darmo w sieci i bez logowania, to już jest odmienny problem prawny. Kolekcja powinna być także łatwa w użyciu (interface, formaty, wyszukiwanie), szczególnie dla użytkowników niepełnosprawnych pod względem ruchowym, czy przy niedowidzeniu. Podstawowe zasady użytkowania (help) powinny być opisane i dostępne online. Biblioteka cyfrowa powinna być widoczna przez wszystkie przeglądarki, działać zarówno pod Windows, jak i Linux czy Mac, trzeba ją testować pod tym kątem. Powinna mieć wersję językową zrozumiałą dla wszystkich. Zob. wersje językowe Śląskiej Biblioteki Cyfrowej: <http://www.sbc.org.pl/dlibra>.

Zasada 5: Dobra kolekcja cyfrowa respektuje wszystkie prawa autorskie.

Zasady respektowania prawa autorskiego powinny być opisane i udostępnione online na stronach AC na poziomie całej kolekcji³ oraz na poziomie obiektu. Należy podać informację o tym, jakie prawa do kolekcji ma instytucja sprawcza i jej partnerzy, co jest elementem ochrony (baza danych, obiekty cyfrowe) i jaki jest jej zakres. Należy zacytować przepisy prawne, które chronią bibliotekę cyfrową jako całość. Należy brać tu pod uwagę fakt, że w Polsce biblioteki cyfrowe często powstają dzięki funduszom publicznym, zatem roszczenie sobie wyłącznych do nich praw i nakładanie restrykcji wobec obywateli jest nieestosowne.

Sprawy prawne są trudne, przepisów prawa trzeba się uczyć, warto mieć bibliotekarza, który będzie się specjalizował w zakresie wyjaśniania praw majątkowych dla naszej kolekcji. Decyzje prawne generują ryzyko, nie zawsze możemy ustalić właściciela praw majątkowych, dlatego zaleca się, by opracować strategię ryzyka, która pozwoli kolekcję szeroko prezentować, ale zminimalizuje ewentualne roszczenia. Należy w takiej strategii brać pod uwagę dobro publiczne oraz ograniczeniu strat majątkowych ewentualnych właścicieli praw. Strategię trzeba udostępnić publicznie, by zachować przejrzystość swoich działań.

Zasada 6: Dobre biblioteki cyfrowe posiadają mechanizmy pozwalające na zbieranie informacji o użyteczności zasobów.

Zasoby cyfrowe muszą być cyklicznie oceniane, należy badać efektywność ich użycia, stan dostępności, to pomaga potem w podejmowaniu decyzji związanych z rozwojem kolekcji. Kryteria, metody i miary w ocenie powinny być dobrane zgodnie z polityką prowadzenia biblioteki cyfrowej. Jeśli mamy do czynienia z kolekcją naukową, to będziemy mierzyli jej wpływ na rozwój edukacji, jeśli z kolekcją regionalną, to wpływ na rozwój turystyki w regionie, itp. Metody badań są znane, trzeba je tak dobrać, by sobie odpowiedzieć na pytanie: *Kto korzysta z BC, czego używa, jak i dlaczego?* Takie badania można zlecić firmie fachowej. Są także dostępne analizatory, technologie, które wspomagają takie obserwacje.

Zasada 7: Dobra biblioteka cyfrowa współdziała z innymi systemami, jest interoperacyjna⁴.

Oznacza to, że oprogramowanie posiada zdolność komunikowania się z innym oprogramowaniem, wymienia metadane, daje możliwości importowania i eksportowania danych, pobierania zasobów, przeglądania przez roboty wyszukiwujące. Opiera się standardy przesyłania danych.

Zasada 8: Dobra kolekcja daje się integrować z zasobami użytkowników i jest wsparciem dla indywidualnej pracy.

Każdy końcowy użytkownik ma swój warsztat pracy, w którym musi często wykorzystać zasoby cyfrowe, jeśli BC jest przyjazna, to wykorzystanie jej obiektów nie jest trudne. Użytkownicy końcowi mogą bardzo wspomagać rozwój bibliotek,

³ Zob. Zasady udostępniania, gromadzenia i zabezpieczania kolekcji w KPBC
<http://kpbc.umk.pl/dlibra/text?id=polityka>.

⁴ Zgodnie z ISO/IEC 2382-01 *Information Technology Vocabulary, Fundamental Terms* (3.3.1.10), definicja interoperacyjności brzmi: "The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units".

ponieważ dostarczają informacji zwrotnej i metodach pracy z zasobami. Niektóre BC dają dodatkowe możliwości użytkownikom, takie jak dodawanie słów kluczowych do obiektów (tagowanie), przysyłanie swoich zasobów cyfrowych do kolekcji, komentowania ich czy eksportowanie opisów bibliograficznych do własnych bibliografii załącznikowych. Oprogramowanie BC powinno zawsze mieć nowe funkcjonalności wspomagające i indywidualizujące prace użytkowników.

Zasada 9: Dobra kolekcja cyfrowa jest zawsze stabilna.

Instytucja zarządzająca kolekcją musi zapewnić nieustanny, wieloletni dostęp do zasobów, co oznacza zabezpieczenie kolekcji na każdym jej poziomie. Stabilizację AC musi zapewnić należy zarówno od strony organizacyjnej, technicznej, jak i finansowej. Jeśli zleca się wykonanie kolekcji cyfrowej instytucji zewnętrznej, należy sobie zapewnić wszelkie prawa do dysponowania tą kolekcją (plikami i metadanymi) w przyszłości. Zapewnienie stabilności, to także zapewnienie przechowywania długoterminowego nie tylko danych, ale i technicznych, merytorycznych oraz administracyjnych informacji, dzięki którym kolekcja istnieje (know-how).

OBIEKTY CYFROWE

Zasada 1: Dobry obiekt zapisywany jest w formacie, który zapewni mu powtórne wykorzystanie.

Istnieje ściśle powiązanie między jakością, z jaką produkuje się obiekt, a jego czytelnością i elastycznością, dzięki którym potem się go powtórnie używa lub migruje między platformami. Przy czym format rozumiemy tu także jako sposób kodowania transmisji bitów. Zalecanych formatów jest tak wiele, jak wiele jest typów dokumentów (text, fotografia, audio, video)⁵. Dla każdego typu dokumentu zaleca się inny format. Trzeba w tym kontekście także widzieć inaczej obiekt cyfrowy uzyskany z digitalizacji a inaczej taki, który urodził się w postaci elektronicznej (born digital). W przypadku tych drugich należy pamiętać, żeby starać się przechowywać te materiały nie w formatach lokalnych, ale takich, które w przyszłości będą mogły być czytelne. Im bardziej popularny format, tym większe prawdopodobieństwo jego łatwej migracji do innego. Pliki matki powstałe w wyniku skanowania zaleca się dziś zabezpieczać w formacie tiff.

Zasada 2: Dobry obiekt cyfrowy musi się dać zachować dla przyszłości. Musi być dostępny przez długi okres⁶.

W tej chwili pracuje się nad dwiema strategiami zachowania obiektów: emulacją i migracją, które są ściśle od siebie zależne. Emulacja polega na bezkolizyjnym przechodzeniu do bardziej współczesnych systemów i urządzeń komputerowych. Migracja jest przenoszeniem obiektów między formatami czy „przepisywaniem” na nowszy format. Strategie te opracowywane są:

- a) dla nośników informacji, które się zmieniają od dysków miękkich przez CD, DVD czy pamięci flash (pen-drive),
- b) dla oprogramowania, które nieustannie się rozwija,
- c) dla formatów i metod kodowania (od SGML do XML), które wybiera się jako standardy w danym momencie.

⁵ Listę formatów zob. w pozycji bibliograficznej nr 1, 8 i 9.

⁶ Szczegóły zobacz online w dokumencie NISO: <http://framework.niso.org/node/19>.

Im bardziej popularne są formaty czy oprogramowanie, tym większa szansa na opracowanie dla nich strategii przetrwania. Kopie matki zaś im bliższe są źródłowego formatu, tym lepiej, wszelkie przetwarzanie powoduje bowiem straty.

Zasada 3: Dobry obiekt jest zrozumiały, użyteczny na zewnątrz poza swoim lokalnym środowiskiem.

Taki obiekt można nazwać interoperacyjnym, wielokrotnie wykorzystywanym, przenoszonym między platformami. Taki obiekt musi być dobrze opisany za pomocą metadanych opisowych (z możliwością przełączania języków) i technicznych oraz musi mieć jasny status, co do możliwości jego wykorzystania. Wszystkie techniczne parametry pozwalające skorzystać z obiektu powinny być dostępne i zrozumiałe dla innych. Obiekt powinien być łatwo odczytywalny przez metawyszukiwarki⁷ czy przenoszony między archiwami cyfrowymi.

Zasada 4: Dobry obiekt powinien mieć stały, unikalny identyfikator, który pozwoli bezbłędnie rozpoznać obiekt i trafić do niego.

Identyfikator⁸ musi mieć taką strukturę, by nie można go było automatycznie duplikować, musi zawierać jakiś kod rozpoznawalny w systemie globalnym, musi być stały i nie może się zmieniać, nawet jeśli obiekt zmieni swoje miejsce w sieci. Musi pozwalać na szybkie odnalezienie dokumentu, najlepiej przez jedno kliknięcie. KPBC ma identyfikator o składni: Identyfikator zasobu: *oai:kpbc.umk.pl:28732*, BCUWr Identyfikator zasobu: *oai:www.bibliotekacyfrowa.pl:21944*.

Zasada 5: Dobry obiekt powinien mieć potwierdzoną autentyczność.

Autentyczność dobrego obiektu może być określona na różne sposoby. Po pierwsze powinien mieć określone pochodzenie, nienaruszoną strukturę i wyraźnie oznaczone kolejne wersje. Obiekt nie może być uszkodzony lub zmieniony w sposób niedozwolony. W metadanych technicznych można określić pochodzenie obiektu (jest taka możliwość w schemacie METS). Do sfotografowanego obiektu można dodać znak wodny określający prawo własności do fizycznego dokumentu, ale nigdy tak, by zakłócał normalne korzystanie z niego. Zobacz znak wodny w KPBC na grafice: *Rzeczka w zimowej szacie* – Stanisława Rolicza (1913-1997) <http://kpbc.umk.pl/dlibra/doccontent?id=27266&dirids=1>. Jeżeli do obiektu upublicznionego online (w zmienionym formacie) dodamy element obcy, np. brakującą kartę, to należy to wyraźnie zaznaczyć w opisie bibliograficznym i podać źródło pochodzenia dodatku. W pliku matce nie wolno dokonywać takich ingerencji, bo nie odzwierciedla to stanu faktycznego.

Zasada 6: Dobry obiekt ma powiązane (wszyte) metadane.

Każdy obiekt powinien być dobrze opisany zarówno od strony merytorycznej (opis bibliograficzny), jak i administracyjnej czy technicznej. Ale szczególnie ważne jest by obiekt miał wszyte (zagnieżdżone, wbudowane w plik) metadane już w momencie

⁷ Metasearch engine. From Wikipedia, the free encyclopedia.

http://en.wikipedia.org/wiki/Metasearch_engine.

⁸ Digital Object Identifier: DOI, zob. Wikipedia.pl: http://pl.wikipedia.org/wiki/DOI_%28identyfikator_cyfrowy%29 lub Frequently Asked Questions about the DOI® System <http://www.doi.org/faq.html#1>.

tworzenia. Takie dane mają np. strony html – są to META tags (znaczniki meta)⁹. Są one istotne z punktu widzenia zarządzania obiektami oraz długotrwałego ich przechowywania. Najważniejsze dane to data utworzenia, pojemność itp.

METADANE

Zasada 1: Dobre metadane¹⁰ mają odniesienie do standardów międzynarodowych związanych z konkretnymi rodzajami danych. Powinny być dobrane do użytkowników tych zasobów oraz zamierzonego, bieżącego i prawdopodobnego przyszłego wykorzystania obiektu cyfrowego.

Nie powinno się stosować własnych rozwiązań, które zamykają nam drogę do międzynarodowej wymiany danych. Powinno się własne dane opisać zgodnie ze standardami przyjętymi powszechnie, by opis był zrozumiany przez każdego użytkownika. Słowniki, tezaury także powinny być dostosowane do naszych użytkowników, ale mieć odniesienie do międzynarodowej rodziny słowników, szczególnie jeśli chodzi o strukturę takich danych. Jeśli kolekcja jest mieszana, to stosuje się różne schematy metadanych, potem można je „opakować” za pomocą kontenera METS. Własne rozwiązania mogą być dodatkiem do standardowych. Polskie biblioteki cyfrowe przy tworzeniu opisów bibliograficznych dla obiektów stosują standard metadanych Dublin Core.

Zasada 2: Dobre metadane wspomagają interoperacyjność.

Oznacza to, że trzeba tak budować metadane, by mogły być łatwo importowane, eksportowane, wykorzystane przy pobieraniu ich do metawyszukiwarek, zgodne z protokołami przesyłania danych (OAI PMH, Z39.50). Dobre metadane muszą być użyteczne w globalnym kontekście. To, co nie wydaje się ważne w lokalnym kontekście, bo jest dla wszystkich jasne, czasem musi być opisane w szerszym. To należy przewidywać. Ważnym jest także pamiętać o tym, że twórcy zasobów cyfrowych mogą współpracować w przyszłości z innymi dostawcami usług, którzy muszą dane standaryzować, implementować do własnych systemów czy łączyć z trzecimi zasobami.

Zasada 3: Dobre metadane stosują kontrolę autorytarną i narzędzia kontroli treści takie jak słowniki kontrolowane, spełniające wymagania użytkowników, co do opisu treści obiektów i łączenia ze sobą obiektów podobnych.

Słowniki haseł wzorcowych są bardzo potrzebne przy tworzeniu wszelkich metadanych, dzięki nim słownictwo jest spójne a opis kontrolowany, co do zgodności z innym takim samym. Klasyfikacje systematyczne porządkują wyszukiwanie danych, grupują je w struktury hierarchiczne, które są bardzo użyteczne w dostępie online. Słowniki dobiera się oczywiście do kolekcji i użytkowników, jacy z nich korzystają.

⁹ Znaczniki meta (ang. meta elements) - zbiór znaczników w sekcji nagłówkowej dokumentu używanych do opisu ich zawartości, stosowany przy tworzeniu stron internetowych w formacie HTML oraz XHTML. Za ich pomocą można przypisać danemu dokumentowi takie wartości, jak adres e-mail, imię i nazwisko autora strony, datę jej powstania słowa kluczowe itp. Można również zadeklarować kodowanie strony. Hasło z Wikipedia.pl: http://pl.wikipedia.org/wiki/Znaczniki_meta.

¹⁰ Metadane – czyli „dane o danych”, ich przykładem są np. klasyczne katalogi biblioteczne są metadanymi. Przy pomocy metadanych opisywane są dokumenty elektroniczne, w szczególności dokumenty dostępne poprzez sieci komputerowe, np. strony World Wide Web, a także dokumenty tworzące nowoczesne biblioteki cyfrowe. Jednym ze standardów metadanych jest Dublin Core Metadata Element Set (DCMES). Hasło z Wikipedia.pl: <http://pl.wikipedia.org/wiki/Metadane>.

Nie ma jednolitych słowników dla wszystkich na świecie, są tylko próby unifikacji niektórych. Wiele polskich bibliotek cyfrowych stosuje (ale niestety nie wszystkie) słownictwo kontrolowane dla nazw osobowych, korporacyjnych, nazw geograficznych oraz haseł przedmiotowych. Wynika to z tego, że eksportują opisy ze swoich katalogów bibliotecznych a te pobierają dane bądź to z NUKAT bądź Biblioteki Narodowej. Niestety nie jest to standard powszechnie stosowany.

Zasada 4: Dobre metadane zawierają jasne określenie warunków i zasad wykorzystywania obiektu cyfrowego.

Przy obiekcie należy umieścić w metadanych opisowy status prawny obiektu, ale nie digitalizatu a utworu intelektualnego czy artystycznego. Pomocny w tym jest poszerzony *Schemat Dublin Core (DC Terms)*, który wprowadza wszystkie istotne elementy opisu dla pola PRAWA. W opisie powinno się zaznaczyć, czy obiekt jest pod ochroną prawa (copyrighted, not copyrighted), czy należy już do domeny publicznej (public domain), kto jest właścicielem praw majątkowych do utworu, na jakiej licencji je udostępnił i na jaki zakres wykorzystania może liczyć użytkownik. Jeśli kolekcja jest jednolita, taka informacja może być zawarta na poziomie opisu kolekcji, jeśli nie, to na poziomie każdego obiektu.

Zasada 5: Dobre metadane techniczne i administracyjne ułatwiają długoterminowe zarządzanie obiektami w zasobach oraz ich zachowanie dla przyszłości.

Metadane techniczne pozwalają na zarządzanie zasobami, dostępem do nich, porządkowaniem ich, strukturalizowaniem, odświeżaniem, opisują pliki, ich format, objętość, wersjonowanie - są one niezbędne na przykład do rekonstrukcji zniszczonego obiektu, przesyłania go, czy przetwarzania. Administracyjne mówią, kto jest odpowiedzialny za obiekty, kto ma do nich dostęp, kto zrobił ich opis i ewentualne musi wnieść poprawki. Wszystkie one muszą być zawsze powiązane z obiektem, gdyż są niezbędne do zarządzania i zabezpieczania kolekcji.

Zasada 6: Dobre rekordy metadanych same w sobie są obiektami, powinny trzymać jakość, powinny być tworzone zgodnie z zasadami tworzenia dobrych obiektów, takimi jak autorytarność, autentyczność, archiwizacja, niezmienność, unikatowość oraz identyfikowalność.

Dlatego powinno się dokładnie opisać, kto, jaka instytucja jest odpowiedzialna za tworzenie metadanych, jakich używa standardów i czy zapewni długoterminowe ich przechowywanie oraz kontrolę i meliorację. Instytucja musi stworzyć warunki do komunikowania się z użytkownikami, w razie gdyby chcieli skomentować jakość metadanych. Często wnoszą oni istotne poprawki do opisów, zauważają braki lub błędy. Wiele schematów metadanych zawiera elementy je opisujące w nagłówkach. Te metadane są niezwykle ważne wobec takich zagrożeń, jak pozycjonowanie stron, fałszerstwa czy spamowanie.

ZAKOŃCZENIE

Nie jest łatwo stworzyć dobry jakościowo produkt, ale o ile łatwiej, kiedy ma się gotowe recepty? Wypracowane standardy i zasady budowania dobrych zasobów cyfrowych są łatwo dostępne. Literatura zagraniczna na ten temat jest ogromna (część z niej można przejrzeć w pozycji bibliograficznej nr 1), wystarczy po nią sięgnąć.

Bibliografia

1. *A Framework of Guidance for Building Good Digital Collections. NISO Recommended Practice*, [Dok. elektr.] Wyd. 3. Framework Working Group with support from the Institute of Museum and Library Services. NISO Baltimore 2007. Tryb dostępu: <http://www.niso.org/framework/>. ISBN (10): 1-880124-74-2.
2. Bednarek-Michalska Bożena, Kujawsko-Pomorska Biblioteka Cyfrowa a standardy, w: Biuletyn EBIB [Dokument elektroniczny]. - Nr 4/2006 (74) kwiecień. - Czasopismo elektroniczne. - [Warszawa] : Stowarzyszenie Bibliotekarzy Polskich KWE, 2006. - Tryb dostępu: <http://www.ebib.info/2006/74/michalska.php>. - Tyt. z pierwszego ekranu. - ISSN 1507-7187.
3. Digital Library Standards and Practices, [dokument elektroniczny] Digital Library Federation Washington DC 2008. [data dostępu 10 marca 2008]. Tryb dostępu: <http://www.diglib.org/standards.htm>.
4. The DOI System. [dokument elektroniczny] The International DOI Foundation 2008. [data dostępu 10 marca 2008]. Tryb dostępu: <http://www.doi.org/>.
5. Metadata. [dokument elektroniczny]. Ukoln 2007, [data dostępu 10 marca 2008]. Tryb dostępu: <http://www.ukoln.ac.uk/metadata/>.
6. Good Practice Guide for Developers of Cultural Heritage Web Services. [dokument elektroniczny]. UKOLN 2004. [data dostępu 10 marca 2008]. Tryb dostępu: <http://www.ukoln.ac.uk/interop-focus/gpg/>.
7. The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materiale. [dokument elektroniczny]. National Initiative for a Networked Cultural Heritage UK. The first edition October 2002. [data dostępu 10 marca 2008]. Tryb dostępu: <http://www.nyu.edu/its/humanities/ninchguide/index.html>.
8. *Digitalizacja piśmiennictwa* pod red. Dariusza Paradowskiego, Biblioteka Narodowa Warszawa 2010. ISBN 978-83-7009-706-6.
9. *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego* pod redakcją i we współpracy Grzegorza Płoszajskiego, Tomasza Kaloty, Dariusza Paradowskiego oraz Kazimierza Schmidta. Politechnika Warszawska, Warszawa 2007. Dostępna także online z Biblioteki Cyfrowej Politechniki Warszawskiej: <http://bcpw.bg.pw.edu.pl/dlibra/docmetadata?id=1262>.